

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 27, 2001, 16:56:13 ; Search time 16.97 seconds

(without alignments)  
505.573 Million cell updates/sec

Title: US-09-830-647-2  
Perfect score: 1206  
Sequence: 1 MNSGAMRHSKHFQGGIQV.....LKKPFYKVEDMSGPVHLM 234

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 3664827 residues  
Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	100.5	8.3	615	YK29_YEAST	P36115 saccharomyc
2	94.5	7.8	1436	DEO3_STAU	O53665 strephilococ
3	91	7.5	908	YV61_YEAST	P42839 saccharomyc
4	91	7.5	994	PPOL_DROME	P36875 saccharomyc
5	90.5	7.5	584	BCAS_HUMAN	O73363 ctosophilin
6	89	7.4	650	DNAR_BURPS	O68191 burkholderi
7	88	7.3	1062	CERU_MOUSE	O61147 mus musculu
8	88	7.3	1214	BRE3_HUMAN	O91144 homo sapien
9	87.5	7.3	885	YDGH_BACSU	P96706 bacillus su
10	86.5	7.2	1462	TOP2_PEA	O24308 plasm sativ
11	85	7.0	602	TP6B_ARCFU	O26605 archaeoglob
12	84.5	7.0	299	SM30_RAT	O03336 rattus norv
13	84	7.0	815	RPGR_HUMAN	O92834 homo sapien
14	84	7.0	819	ADVL_HUMAN	O75366 homo sapien
15	84	7.0	1059	CERU_RAT	P13635 rattus norv
16	83	6.9	650	DNAR_BURPS	P42373 burkholderi
17	83	6.9	1176	YOH8_YEAST	Q08236 saccharomyc
18	82.5	6.8	434	GATA_METUA	O55560 methanococc
19	82.5	6.8	482	ODP2_YEAST	P14695 saccharomyc
20	82.5	6.8	592	ABP1_YEAST	P15891 saccharomyc
21	82.5	6.8	1225	SMCL_YEAST	P33908 saccharomyc
22	82	6.8	297	RRPP_RABYP	P06747 rabies viru
23	82	6.8	304	MMGL_MOUSE	P43500 mus musculu
24	82	6.8	320	YVJD_BACSU	P45864 bacillus su
25	82	6.8	322	RLUC_HAEIN	P44433 haemophilus
26	82	6.8	686	MYBB_CHICK	O03237 gallus gall
27	81.5	6.8	299	SM30_MOUSE	O64374 mus musculu
28	81.5	6.8	379	VLEF1_NPVAC	O06687 autographa
29	81.5	6.8	706	YK70_YEAST	P36166 saccharomyc
30	81.5	6.8	1131	AC15_MOUSE	P33601 mus musculu
31	81	6.7	622	PPID_HAEIN	P44092 haemophilus
32	81	6.7	700	GUNA_PAEIA	P29719 paenibacill
33	81	6.7	1023	PIB4_BOVIN	Q07722 bos taurus

## ALIGNMENTS

RESULT	ID	YK29_YEAST	STANDARD:	PRT:	615 AA.
34	81	6.7	1478	1	BCK1_YEAST
35	81	6.7	4451	1	GRSB_BACAR
36	80.5	6.7	224	1	Y945_METUA
37	80.5	6.7	388	1	VE2_HPV27
38	80.5	6.7	434	1	YK12_YEAST
39	80.5	6.7	673	1	TF2B_METUA
40	80.5	6.7	906	1	CERC_MOUSE
41	80.5	6.7	1523	1	SON_HUMAN
42	80	6.6	328	1	SW16_SCHPO
43	80	6.6	726	1	BPL1_HUMAN
44	80	6.6	1022	1	PIB4_HUMAN
45	79.5	6.6	312	1	KIME_METUA

  

001389 saccharomyc	P14688 b granulicidi
O58355 methanococc	P36789 human papill
P36080 saccharomyc	O58192 methanococc
P49452 mus musculu	P18583 homo sapien
P40381 schizosacch	P50747 h biotin-p
O15147 homo sapien	O58487 methanococc

  

3	SCAMRHSKHFQGGIOVNEKNRPS-----LKSLEKTDNRPEKSCPKPLMGKV	50
150	NGULGHSRSGSRRNG--SNESLTPGQRTPPDKRSQENLTSFSSGRRSSSSSHEPATTND	206
51	FYLDPSVTYSKLDKIDKIDGRIYE-----EFLSKDISYLSLNKKEAKFAQT	98
207	SNKALPKRRPSPPLQSLVSGQLHENEHLISISIDRSRLNPDTSIDVSNRSQTSLSOT	266
99	LKRISVPPESEAYTAETSPRPHSDGSSFKS--PTTVLSNGKILYER-----	145
267	INQLSTCESEPIASNTTYYT-TSNGSGLPMLVPPYSSDMKKKKLYNFKRKRVGSKPK	325
146	-----AIKDHFIPS-----NSLSNALSGVKILHDDIRYYTEQKK	183
326	HLSSQYEMASSEELGQHGOQPSMRKTTILKRTSVSTNMFNDHASSLHGKNIKYK-NPS	384

QY 184 KELYLKSSSTSVRDGKRVSGAOKRTGTGLKPKPEYKVEDMSQSPAVH 232  
DB 385 NDTVDVDDTDDSDSDONODALTRPKRRDRIRK--KIRNSANKTAHH 430

RESULT 2  
ID DPO3\_STAAU STANDARD: PRT: 1436 AA.  
AC Q53665: 057110: 09F1T9;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 20-AUG-2001 (Rel. 40, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE DNA POLYMERASE III POLC-TYPE (EC 2.7.7.7) (POLIII).  
GN POLC.  
OS Staphylococcus aureus.  
OC Bacteria; Firmicutes; Bacilli; Clostridium group.  
OC Bacillus/Staphylococcus group; Staphylococcus.  
NCBI\_TaxID=1280.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN:128;  
RX MEDLINE=96084951; PubMed=7489915;  
RA Pacetti D.F., Barnes M.H., Li D., Brown N.C.;  
RT \*Characterization and overexpression of the gene encoding  
RT Staphylococcus aureus DNA polymerase III.;  
RN Gene 165:51-56(1995).  
[2]  
RP SEQUENCE FROM N.A.  
RA Inoue R., Kaito C., Tanabe M., Kamura K., Akimitsu N., Sekimizu K.;  
RT \*Genetic identification of two distinct DNA polymerases, Dnae and  
RT PolC, essential for chromosomal DNA replication in Staphylococcus  
RT aureus.";  
RN Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: REQUIRED FOR REPLICATIVE DNA SYNTHESIS. THIS DNA  
CC POLYMERASE ALSO EXHIBITS 3' TO 5' EXONUCLEASE ACTIVITY.  
CC -1- CATALYTIC ACTIVITY: N DROXYNUCLEOSIDE TRIPHOSPHATE -  
CC N PYROPHOSPHATE + DNA(N).  
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-C FAMILY. POLC  
CC SUBFAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: D86727; BAA13160.1;  
DR EMBL: Z48003; CAAB8043.1;  
DR EMBL: AB053353; BAB20885.1;  
DR InterPro: IPR003141; PNP.N.  
DR InterPro: IPR000520; Exonuclease.  
DR Pfam: PF00929; Exonuclease; 1.  
DR Pfam: PF02231; PNP.N. 1.  
DR SMART: SM00479; EXOIII; 1.  
DR SMART: SM00481; POLITFAC; 1.  
DR Transferrase: DNA-directed DNA polymerase; DNA replication; Hydrolase;  
KM Nuclease; Exonuclease.  
KW DOMAIN 420 584  
FT DOMAIN 327 208 NE -> KO (TN REF. 1).  
FT CONFLICT 352 352 T -> K (TN REF. 1).  
FT CONFLICT 1030 1049 MISSING (TN REF. 1); CAAB8043).  
FT CONFLICT 1035 1035 R -> A (TN REF. 1).  
FT CONFLICT 1147 1153 EFGTGFV -> NSDRIR (TN REF. 1).  
FT CONFLICT 1260 1260 M -> I (TN REF. 1).  
FT CONFLICT 1409 1409 S -> Y (TN REF. 1).  
SQ SEQUENCE 1436 AA: 162459 MW: 2F70E034C0FBF723 CRC64;

Query Match 7.8%; Score 94.5; DB 1; Length 1436;

Best Local Similarity 25.1%; Pred. No. 9.4;  
Matches 57; Conservative 32; Mismatches 87; Indels 51; Gaps 12;

QY 18 IQVKNENKRNPSLAKTNDNR- ESKCKPLMKVFYLDPSVTISEKLODKDGRVE 76  
DB 805 VPIDELCTPRMEGANEELRELESTANAKLGE---DLQIVT-DLEKEKSIQN-- 857  
QY 77 EFLSKDISYLIISN--KKEAKPAOTLGRISFPVPSASATYAEFT--SRPSSH----- 123  
DB 858 --GEAVIYLIISORLVKRSGLDGYVSGSGSVGSFYATWTEITTEYNPLPPHYICPMCKT 914  
QY 124 -----DGSSEKSPDYVCLSRKLLVERAKIDHDFIPSNISLSNLSMGVYKLIHDDI 175  
DB 915 SEFPNDGVSQGFDPDKTCFCGAPL----IKGODIFETFELEFK--GDKVPDI-DL 966  
QY 176 RYVTEOKKELLYLKSSYSVADGKRV-----GSGAOKRTYGRLK 216  
DB 967 NFSGEYQPN-----AHNTKVLFGEDKVFAGTIGTVAEKTAFGYVK 1008

RESULT 3  
ID YN6L\_YEAST STANDARD: PRT: 908 AA.  
AC P42839;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE HYPOTHETICAL 102.5 KDA PROTEIN IN KRE1-HXT14 INTERGENIC REGION.  
GN YNL321W OR N0339.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetaceae; Saccharomycetaceae; Saccharomycetes.  
NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288C / FY1676;  
RX MEDLINE=96076632; PubMed=7502583;  
RA Maltahl M., Nicaud J.-M., Levesque H., Galliard C.;  
RT \*Sequencing analysis of a 24.7 kb fragment of yeast chromosome XIV  
RT identifies six known genes, a new member of the hexose transporter  
RT family and ten new open reading frames.";  
RN Yeast 11:1077-1085(1995).  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).  
CC -1- SIMILARITY: NO S. POMBE SPAC521.04C.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: Z46259; CAAB6376.1;  
DR EMBL: Z71597; CAAB6252.1;  
DR SGD: S0005265; YNL321W.  
KW Hypothetical protein; Transmembrane.  
FT TRANSMEM 245 265 POTENTIAL.  
FT TRANSMEM 409 429 POTENTIAL.  
FT TRANSMEM 495 515 POTENTIAL.  
FT TRANSMEM 531 551 POTENTIAL.  
FT TRANSMEM 561 581 POTENTIAL.  
FT TRANSMEM 588 608 POTENTIAL.  
FT TRANSMEM 627 647 POTENTIAL.  
FT TRANSMEM 627 707 POTENTIAL.  
FT TRANSMEM 747 767 POTENTIAL.  
FT TRANSMEM 784 804 POTENTIAL.  
FT TRANSMEM 817 837 POTENTIAL.  
FT TRANSMEM 852 872 POTENTIAL.  
FT TRANSMEM 886 906 POTENTIAL.  
SQ SEQUENCE 908 AA: 102498 MW: 180E0BE077DC111E CRC64;



KM ADP-ribosylation: zinc-finger; zinc; Alternative splicing.  
 FT DNA\_BIND 1 367  
 FT DOMAIN 368 507  
 FT DOMAIN 380 454  
 FT DOMAIN 508 594  
 FT ZN\_FING 19 54  
 FT ZN\_FING 123 161  
 FT DOMAIN 208 210  
 FT DOMAIN 223 228  
 FT ACT\_SITE 875 875  
 FT VARSPPLIC 376 564  
 SQ SEQUENCE 994 AA; 113791 MW; ACAB5A270DD29E08 CRC64;

Query Match 7.5%; Score 91; DB 1; Length 994;  
 Best Local Similarity 21.2%; Pred. No. 11;  
 Matches 62; Conservative 46; Mismatches 80; Indels 104; Gaps 14;

QY 21 KNEKNRPSLSTKTDN--RPEKSKCPPLMGKVFYLDLPSTVTS---EKLOKIDKLGGR 74  
 DB 356 KSTFSKSLKTKNNKNDLVRLPTLRSP--PLYNKFSITGLKNQKHEKRLKRIENLGGK 412  
 QY 75 VEEFLSKDISYLSNKKR-----AKFAQTIG-RISPV----- 105  
 DB 413 FEVKISENTIATISTELEIQKSTRMKFAEELGIHVPLEFDEVEADTEGAIKYINSTC 472  
 QY 106 -----PSPESAIVTAETSPHPSHD-----GSSKSPDTVCLSRGKLYEKAIDHDF 152  
 DB 473 ICSWGDPRKSRIPKRTKSLNSNSITKSWPVSRTFKYKVDGLAVDPDSGLEDIA--HY 529  
 QY 153 IFSNLSLMSWGVKILIHDDIR-----IY-----IEOKKELLYLKSSSVADGSKRY 203  
 DB 530 VDSNNKYS-----VVLGLDIDQKKNKSYKVQOLKADKKEKWFPSGRI---GTNI 579  
 QY 204 GSG-----AQKT-----RTGRLKKPKYKED 224  
 DB 580 GNSKLEPDTESAKRNFKEIYADKRGNEEDORDNFVKRGRMFIPIEIOYD 631

RESULT 5  
 BCAS\_HUMAN STANDARD; PRT; 584 AA.  
 ID 075363;  
 DT 20-AUG-2001 (Rel. 40, Created)  
 DT 20-AUG-2001 (Rel. 40, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE BREAST CARCINOMA AMPLIFIED SEQUENCE 1 (NOVEL AMPLIFIED IN BREAST  
 DE CANCER 1) (AMPLIFIED AND OVEREXPRESSED IN BREAST CANCER).  
 DE BCAS1 OR NABCI OR AIBCI.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 NCBI\_TaxID:9606;  
 RN [1]  
 RP MEDLINE:98337975; PubMed:9671742;  
 RA Collins C., Komens J.W., Kowbel D., Godfrey T., Tanner M.,  
 RA Huang S.-I., Polikoff D., Nonet G., Cochran J., Myambo K., Jay K.E.,  
 RA Giordano J., Cloutier T., Kuo W.-L., Taswen P., Dalziel S.,  
 RA Palazolo M., Martin C., Ericsson C., Plunket D., Albertson D.,  
 RA Li W.-B., Gray J.W.;  
 RT "Positional cloning of ZNF217 and NABCI: genes amplified at 20q13.2  
 RT and overexpressed in breast carcinoma."  
 RL Proc. Natl. Acad. Sci. U.S.A. 95:8703-8708(1998).  
 CC -1- TISSUE SPECIFICITY: OVEREXPRESSED IN MOST BREAST CANCER CELL  
 CC LINES.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AF041260; AAC39896.1; -  
 DR MIM: 602968; -  
 SQ SEQUENCE 584 AA; 61719 MW; 2290CC0B37129944 CRC64;

Query Match 7.5%; Score 90.5; DB 1; Length 584;  
 Best Local Similarity 26.3%; Pred. No. 6;  
 Matches 36; Conservative 22; Mismatches 50; Indels 29; Gaps 6;

QY 23 EKNRPSLSTKTDNRPEKSKCPPLMGKVFYLDLPSTVTSFKIDKIDKLGGRVEEFLSKD 82  
 DB 282 ENNSNIMFFETLVSPKKAETK-----KDPEDTGAESPTTSAD 320  
 QY 83 IYLSNKKR--AKFAQTIGRISPVSPESATTAETSPHPSHDSSGSSFKSPDTVCLSRGK 140  
 DB 321 LK--SDKANFTSGETGAGAKNSKGCNP--SGHTQSVITPEPAKGTREKSGPT-SLPLCK 375  
 QY 141 LVEKAIDHDFIPSSNS 157  
 DB 376 LEWKSXKE-DSVPTGA 391

RESULT 6  
 DNAK\_BURPS STANDARD; PRT; 650 AA.  
 ID 068191;  
 DT 20-AUG-2001 (Rel. 40, Created)  
 DT 20-AUG-2001 (Rel. 40, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE CHAPERONE PROTEIN DNAK (HEAT SHOCK PROTEIN 70) (HEAT SHOCK 70 KDA  
 DE PROTEIN) (HSP70).  
 GN DNAK.  
 OS Burkholderia pseudomallei (pseudomonas pseudomallei).  
 OC Bacteria; Proteobacteria; beta subdivision; Burkholderia group;  
 OC Burkholderia.  
 NCBI\_TaxID:28450;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ATCC 23343;  
 RA See L.H., Yap E.H., Yap E.P.H.;  
 RT "Isolation and sequencing of the heat shock protein 70 (hsp70/dnak)  
 RT gene in Burkholderia pseudomallei."  
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: ACTS AS A CHAPERONE (BY SIMILARITY).  
 CC -1- INDUCTION: BY STRESS CONDITIONS E.G. HEAT SHOCK (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AF016711; NC015473.1; -  
 DR HSP; F04475; ZBRP.  
 DR Interpro: IPR001023; HSP70.  
 DR Pfam: PF00012; HSP70; 1.  
 DR PRINTS: PR00301; HEATSHOCK70.  
 DR PROSITE: PS00297; HSP70\_1; 1.  
 DR PROSITE: PS00329; HSP70\_2; 1.  
 DR PROSITE: PS01036; HSP70\_3; 1.  
 DR CHAPERONE; ATP-binding; Heat shock.  
 KW SEQUENCE 650 AA; 69735 MW; BE46330B6DE174D0 CRC64;

Query Match 7.4%; Score 89; DB 1; Length 650;  
 Best Local Similarity 24.0%; Pred. No. 9.2;  
 Matches 50; Conservative 36; Mismatches 74; Indels 48; Gaps 12;

```

QY 17 GIOVNEKRNPSLSKLT-----KTDNREPKSKCKPLMGKVFILDLPSVT-ISEKLODKID 70
DB 162 GLEVKRIINEPRAALAEFLDAEKGDRKIAVYDLGGTFVDYSITIELADVGEMOEVLIS 221
QY 71 -----LGGREVEFLSKDISYLSLN-KKE-----AKFAOTLGRISFVSP-----ESAAY 113
DB 222 TNGDFIFLGG--EDFPQRITDIYIGEFKKEGVDLKSDVALQRIKRAAKAKRIEISSSQ 279
QY 114 AETTSBPHSHDSSSKSPVTVCLSRGKL-----LYEKAIKHDFIPSNLSISLALSGV 167
DB 280 TEINIPYITADASGRHLN-LKVTRAKLEALVEDIVERTIE-----PCRTAIKDA---GV 330
QY 168 KILHIDDIRY-----IEOKKELY 187
DB 331 KVSDDIDVILVGGQTRMPKVQENKEEF 358

RESULT 7
CERU_MOUSE STANDARD: PRT: 1062 AA.
AC 061147.
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE CERULOPLASMIN PRECURSOR (EC 1.16.3.1) (FERROXIDASE).
CN CP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Klomp L.W.J., Farhangrazi Z.S., Choi D.W., Gitlin J.D.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP TISSUE SPECIFICITY.
RX MEDLINE=96294736; PubMed=86690795;
RA Klomp L.W.J., Farhangrazi Z.S., Dugan L.L., Gitlin J.D.;
RL J. Clin. Invest. 98:207-215(1996).
CC -1- FUNCTION: CERULOPLASMIN IS A BLUE, COPPER-BINDING (6-7 ATOMS PER MOLECULE) GLYCOPROTEIN FOUND IN PLASMA. FOUR POSSIBLE FUNCTIONS ARE FERROXIDASE ACTIVITY, AMINE OXIDASE ACTIVITY, COPPER TRANSPORT AND HOMEOSTASIS, AND SUPEROXIDE DISMUTASE ACTIVITY.
CC -1- CATALYTIC ACTIVITY: 4 FE(2+) + 4 H(+) + O(2) -> 4 FE(3+) + 2 H(2)O.
CC -1- COFACTOR: BINDS 6 CU-IONS PER MOLECULE. THIS PROTEIN BELONGS TO THE MULTICOPPER OXIDASES WHICH CONTAIN THREE DISTINCT CU CENTERS KNOWN AS TYPE 1 OR BLUE, TYPE 2 OR NORMAL, AND TYPE 3 OR COUPLED BINDING.
CC -1- TISSUE SPECIFICITY: MANY TISSUES, INCLUDING LIVER, EYE AND BRAIN.
CC -1- SIMILARITY: CONTAINS 3 F5/8 TYPE A DOMAINS; EACH IS COMPOSED OF 2 PLASTOCYANIN-LIKE REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation- the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U49430; AAB07996.1; -.
CC DR HSSP: P00450; IKCM.
CC DR MGD: MGI:88476; CP.
CC DR InterPro: IPR001117; Cu-oxidase.
CC DR InterPro: IPR002355; Multicu_oxidase2.
CC DR Pfam: PF00394; Cu-oxidase; 3.
CC DR PROSITE: PS00079; MULTICOPPER_OXIDASE1; 3.
CC DR PROSITE: PS00080; MULTICOPPER_OXIDASE2; 1.
CC DR Oxidoreductase; Copper; Metal-binding; Glycoprotein; Plasma; Repeat;
KW Signal.
FT SIGNAL 1 19 BY SIMILARITY.

```

```

FT CHAIN 20 1062 CERULOPLASMIN.
FT DOMAIN 20 356 F5/8 TYPE A 1.
FT DOMAIN 20 189 PLASTOCYANIN-LIKE 1.
FT DOMAIN 208 199 PLASTOCYANIN-LIKE 2.
FT DOMAIN 369 713 F5/8 TYPE A 2.
FT DOMAIN 369 713 PLASTOCYANIN-LIKE 3.
FT DOMAIN 565 713 PLASTOCYANIN-LIKE 4.
FT DOMAIN 725 1057 F5/8 TYPE A 3.
FT DOMAIN 725 896 PLASTOCYANIN-LIKE 5.
FT DOMAIN 904 1057 PLASTOCYANIN-LIKE 6.
FT DISULFID 173 199 BY SIMILARITY.
FT DISULFID 275 356 BY SIMILARITY.
FT DISULFID 529 555 BY SIMILARITY.
FT DISULFID 632 713 BY SIMILARITY.
FT DISULFID 870 896 BY SIMILARITY.
FT METAL 120 120 COPPER (TYPE 2) (BY SIMILARITY).
FT METAL 122 122 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 179 179 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 181 181 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 990 990 COPPER (TYPE 1) (BY SIMILARITY).
FT METAL 993 993 COPPER (TYPE 2) (BY SIMILARITY).
FT METAL 995 995 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 1035 1035 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 1036 1036 COPPER (TYPE 1) (BY SIMILARITY).
FT METAL 1037 1037 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 1041 1041 COPPER (TYPE 1) (BY SIMILARITY).
FT METAL 1046 1046 COPPER (TYPE 1) (BY SIMILARITY).
FT CARBOHYD 138 138 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 226 226 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 583 583 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 625 625 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 757 757 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 922 922 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1062 AA; 121159 MW; F3F52ED09A238F16 CRC64;

Query Match 7.3%; Score 88; DB 1; Length 1062;
Best Local Similarity 21.2%; Pred. No. 21;
Matches 45; Conservative 30; Mismatches 81; Indels 56; Gaps 10;

QY 10 SKGHFGGIQVKNKRNPSLSKLTNDNREPKSKCKPLW-----GKVEYLDLPSVTISEKLO 65
DB 52 SNFYLQNGPRIGRIKRYKKALYFETDTGTFSTKIDKPAWILGFLGPIV-----KAEVEDKYV 106
QY 66 KDKIKDLGGRVEFLSKDISYLSLNKKEAKFAOTLGRISP-----VPSPESAT 112
DB 107 VHLKNLASRIYTFPHAGVY---TKREY-----GAVYPDNTTDFORADKVLPGGOQYV 157
QY 113 TARTTSPHP-----SH-----DGSSFSPTVCLSRGKLVEKAIK-DHDF 152
DB 158 VILHANEPSRGSDSNKCVTRIYHSHVDAPKDIALSGLIGPLILC-KKGSLYKKEKKNIDQEF 216
QY 153 IPSNLSISNALSGVKRLIHLIDIRYIYEOKK 184
DB 217 VLMFSYVDENLSWYLE-----DNLIKFCSEPEK 244

RESULT 8
BREF3_HUMAN STANDARD: PRT: 1214 AA.
AC 090UD4.
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE BROMODOMAIN AND PHD FINGER-CONTAINING PROTEIN 3 (FRAGMENT).
CN BRP3 OR KIAA1286.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
FT TISSUE=Brain;

```

```

RX MEDLINE=20039619; PubMed=10574462;
RA Nagase T., Ishikawa K.-I., Kikuno R., Hirose M., Nomura N.,
RA Ohara O.;
RT Prediction of the coding sequences of unidentified human genes. XV.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.
RL DNA Res. 6:337-345(1999).
CC -1- SIMILARITY: CONTAINS 1 BROMODOMAIN.
CC -1- SIMILARITY: CONTAINS 1 PWWP DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 PWWP DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AB033112; BAA8600.1;
DR Interpro: IPR001487; Bromodomain.
DR Interpro: IPR001965; PHD.
DR Interpro: IPR000313; PWWP.
DR Pfam: PF00439; bromodomain.1.
DR Pfam: PF00628; PHD.1.
DR Pfam: PF00825; PWWP.1.
DR PRINTS: PRO0503; BROMODOMAIN.
DR SMART: SM00337; BROMO.1.
DR SMART: SM00339; PHD.2.
DR SMART: SM00363; PWWP.1.
DR PROSITE: PS00013; BROMODOMAIN_1; FALSE_NEG.
DR PROSITE: PS00014; BROMODOMAIN_2; 1.
DR PROSITE: PS00812; PWWP.1.
DR KEGG: K11611; Bromodomain.
FT ZN-FING 221 271 PHD-TYPE.
FT DOMAIN 415 441 GLU-RICH.
FT DOMAIN 615 685 BROMODOMAIN.
FT DOMAIN 1085 1168 PWWP.
FT SEQUENCE 1214 AA; 136598 MW; CA490810622109CD CRC64;
SQ

```

Query Match 7.3%; Score 88; DB 1; Length 1214;  
 Best Local Similarity 21.5%; Pred. No. 25;  
 Matches 51; Conservative 40; Mismatches 80; Indels 66; Gaps 11;

```

OY 14 FGGIOGVKNKRNPSLSKLTQNRPEKSKCPPLMGKVFYLDLPSVITSEKLOKDK-DLG 72
DB 6 FPGAMRPRKRSQNMGRSPS-PYSLKCSPTRETLTY-----AQAORIVEDVID 55
OY 73 GRV-----EPLSKDISYLSKKNKKAQAQTLGRISPPSPESAVYAETTS 118
DB 56 GRLHRTSYDPLKRTTEDELTAODITECNSKNSKESQOPFGK-SKRPSSKCK-KKESCS 113
OY 119 PPSHDSGSSFSPP-----DYVCLSRGKLT-----VEKAKDIDFIPSSILSNLS 164
DB 114 KHAS--GTSFHLPPQSPRWVSDGIQPEAPLPAAYRYRIKPPEDDAVEYDMDEDLA 171
OY 165 W-----GKYLILHIDIRYIEQKKRELYLLKSSSTVYDGGKRYGSAQKT 210
DB 172 WLDVNYEKRVYDGHSLVSAFTFELVDRLEKESYLESR-----SGAQS 216

```

RESULT 9  
 YDGH\_BACSU STANDARD; PRT; 885 AA.

```

AC P96706;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PUTATIVE MEMBRANE PROTEIN YDGH.
GN YDGH.
OS Bacillus subtilis.

```

```

OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Kasahara Y., Nakai S., Lee S., Sadate Y., Ogasawara N.;
RT "A 148 kbp sequence of the region between 35 and 47 degree of the
RT Bacillus subtilis genome."
RL Submitted (MAR-1997) to the EMBL/GenBank/DDay databases.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE MAPL FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AB001488; BAA19398.1;
DR EMBL: 299106; CAB12372.1;
DR EMBL: 299107; CAB12384.1;
DR Subtilist; BG12175; ydgh.
DR Interpro: IPR000731; HMGCR_patched_STM.
DR PROSITE: PS0156; SSD.1.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 9 29 POTENTIAL.
FT TRANSMEM 181 201 POTENTIAL.
FT TRANSMEM 202 222 POTENTIAL.
FT TRANSMEM 227 247 POTENTIAL.
FT TRANSMEM 278 298 POTENTIAL.
FT TRANSMEM 304 324 POTENTIAL.
FT TRANSMEM 354 374 POTENTIAL.
FT TRANSMEM 716 736 POTENTIAL.
FT TRANSMEM 740 760 POTENTIAL.
FT TRANSMEM 772 792 POTENTIAL.
FT TRANSMEM 817 837 POTENTIAL.
FT TRANSMEM 847 867 POTENTIAL.
FT SEQUENCE 885 AA; 95488 MW; BDCBE161911D75B CRC64;
SQ

```

Query Match 7.3%; Score 87.5; DB 1; Length 885;  
 Best Local Similarity 23.3%; Pred. No. 18;  
 Matches 47; Conservative 30; Mismatches 74; Indels 51; Gaps 8;

```

OY 9 HSKGHFQGIQVKNKRNPSLSKLTQDN-----REKSKCPPLMGKVFYLD 54
DB 34 NQKGQQLPADVNSERANALIKQAGDNNSISVFTLDMALKKETEMQLIIDTKKD 93
OY 55 -LPVYITSEKLOKDKLQGRVPEFLSKD-----ISYLSNKKKKAQOTGRISPPV 106
DB 94 GAEVYISPLSAKEKVK-----QLMSKQKTYLMPYITIGSDKAKKINDEIYQ--VP 145
OY 107 SPESATYATFT-----SPHPSHDSGSSFSPPDYVCLSRGKLTVEKAKIDHDFIP----- 154
DB 146 DDLATITITVQSLINDPFAHSEBQDKKTVITVCLIGLILIVRSVVTPIYVVGFS 205
OY 155 ---SNSILSNLSMGKYLILHID 173
DB 206 YLISOSIL-----GILVYNYD 221

```

RESULT 10  
 TOP2\_PEA STANDARD; PRT; 1462 AA.

```

AC Q24306;
DT 13-DEC-1998 (Rel. 37, Created)
DT 13-DEC-1998 (Rel. 37, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE DNA TOPOLISOMERASE II (EC 5.99.1.3).
GN TOP2 OR TOP11.
OS

```

05 Pisum sativum (Garden pea).  
 0C Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 0C Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 0C eustosids I; Fabales; Fabaceae; Papilionoideae; Viciae; Pisum.  
 0X NCBI\_TaxID=3888;  
 01  
 RN SEQUENCE FROM N.A.  
 RP TISSUE-Leaf;  
 RC Reddy M.K., Nair S., Tewari K.K.;  
 RA Submitted (Aug-1997) to the EMBL/GenBank/DBJ databases.  
 RL -1- FUNCTION: CONTROL OF TOPOLOGICAL STATES OF DNA BY TRANSIENT  
 CC BRANKAGE AND SUBSEQUENT REJOINING OF DNA STRANDS. TOPOISOMERASE II  
 CC MAKES DOUBLE-STRAND BREAKS.  
 CC -1- CATALYTIC ACTIVITY: ATP-DEPENDENT BREAKAGE, PASSAGE AND REJOINING  
 CC OF DOUBLE-STRANDED DNA.  
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).  
 CC -1- MISCELLANEOUS: EUKARYOTIC TOPOISOMERASE I AND II CAN RELAX BOTH  
 CC RELAX ONLY NEGATIVE SUPERCOILS.  
 CC -1- SIMILARITY: BELONGS TO THE TYPE II TOPOISOMERASE FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 CC EMBL: Y14559; CAA74891.1;  
 DR InterPro: IPR000947; CRPA\_NFYB.  
 DR InterPro: IPR001241; DNA\_topoisom.  
 DR InterPro: IPR002205; DNA\_topoisom.  
 DR InterPro: IPR003594; HATPase\_c.  
 DR Pfam: PF00204; DNA\_topoisomI; 1.  
 DR Pfam: PF00521; DNA\_topoisomI; 2.  
 DR Pfam: PF02518; HATPase\_c; 1.  
 DR PRINTS: PR00418; TP12FAMILY.  
 DR PRINTS: PR00615; CCATSUBUNITA.  
 DR PRINTS: PR01158; TOPISMRASE1.  
 DR PRODOM: PD000616; DNA\_topoisomI; 1.  
 DR SMART: SM00433; TOP2c; 1.  
 DR SMART: SM00434; TOP2c; 1.  
 DR PROSITE: PS00177; TOPOISOMERASE\_II; 1.  
 DR Isomerase; Topoisomerase; DNA-binding; ATP-binding.  
 KW NP\_BIND 149 154 ATP (POTENTIAL).  
 FT ACT\_SITE 761 761 DNA\_CLEAVAGE (BY SIMILARITY).  
 SQ SEQUENCE 1462 AA; 164205 MW; D9212C54A0F0F8B2E CRC64;

Query Match 7.2%; Score 86.5; DB 1; Length 1462;  
 Best Local Similarity 20.7%; Pred. No. 42;  
 Matches 53; Conservative 30; Mismatches 88; Indels 85; Gaps 12;

0Y 21 KNEKRRPSLAKTQNRPEKSKCP-----LMGKXYFLDPSVTI----- 60  
 DB 1044 KGRKRPVAGANNDDSEQDAEPETASQSVSECATWGD--YDDLISLPIGLITLLESV 1101  
 0Y 61 ----SEKLOKD-----IKDLGRVEEFLSK-----DISYLSNKKREAK--- 94  
 DB 1102 OKLDEKTEKEVEYELSTPTTSLIMAKL---DEFKKKDELDELKVAEDDKRRASQSS 1157  
 0Y 95 ----FAOTLGRISVPSPESAVTAETTSPPHSIDGSSFK-----SPDTVCISRGTL 141  
 DB 1158 KRANGFASPAKPPQPRKN---TKKAKSVPEPENNSSMEIENAVEAKPAEYAKPKGKA 1214  
 0Y 142 LVEKAIKHDIFPSNSILSNALSMGWKILHIDIRYITQKKKELYLLKSSSTVADGK 201  
 DB 1215 AKKNIOE--PEDDIS-----LOERLAANISSSEKSGSOAMEEVEVOOAKAK 1262  
 0Y 202 RVSGSGAOKRTGRGLK 217  
 DB 1263 K-----QNNKRGAKK 1273

RESULT 11  
 TP6B\_ARCFU STANDARD; PRT; 602 AA.  
 ID TP6B\_ARCFU  
 AC 029605;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE TYPE II DNA TOPOISOMERASE VI SUBUNIT B (EC 5.99.1.3).  
 GN TP6B OR AF0652.  
 OS Archaeoglobus fulgidus.  
 OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;  
 CC Archaeoglobus  
 0X NCBI\_TaxID=2234;  
 01  
 RN SEQUENCE FROM N.A.  
 RP STRAIN-VC-16 / DSM 4304 / ATCC 49558;  
 RC MEDLINE=98049343; PubMed=9389475;  
 RX Klek H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,  
 RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,  
 RA Richardson D.B., Kellavagge A.R., Graham D.E., Kyriades N.C.,  
 RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,  
 RA Kirdness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,  
 RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,  
 RA Overbeek R., Cocayne J.D., Weisman J.F., McDonald L., Utterback T.,  
 RA Colton M.D., Spriggs T., Artlich P., Kaine B.P., Sykes S.M.,  
 RA Sadow P.W., D'Andrea K.P., Bowman C., Fujil C., Garland S.A.,  
 RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,  
 RA Venter J.C.;  
 RT "The complete genome sequence of the hyperthermophilic, sulphate-  
 RT reducing archaeon Archaeoglobus fulgidus.";  
 RL Nature 390:364-370(1997).  
 CC -1- FUNCTION: RELAXES BOTH POSITIVE AND NEGATIVE SUPERURNS AND  
 CC EXHIBITS A STRONG DECATENASE ACTIVITY. THE B SUBUNIT BINDS ATP (BY  
 CC SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: ATP-DEPENDENT BRANKAGE, PASSAGE AND REJOINING  
 CC OF DOUBLE-STRANDED DNA.  
 CC -1- SUBUNIT: HETEROTETRAMER OF TWO SUBUNITS A AND TWO SUBUNITS B (BY  
 CC SIMILARITY).  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 CC EMBL: AE001060; AAB90588.1; -;  
 DR TIGR: AF0652; -;  
 DR InterPro: IPR000410; Bcrl1\_sensor.  
 DR InterPro: IPR003594; HATPase\_c.  
 DR Pfam: PF02518; HATPase\_c; 1.  
 DR SMART: SM00387; HATPase\_c; 1.  
 KW Isomerase; Topoisomerase; DNA-binding; ATP-binding; Complete proteome.  
 SQ SEQUENCE 602 AA; 67399 MW; FBA2111CF4280DC7 CRC64;

Query Match 7.0%; Score 85; DB 1; Length 602;  
 Best Local Similarity 22.4%; Pred. No. 17;  
 Matches 47; Conservative 36; Mismatches 85; Indels 42; Gaps 9;

0Y 20 VKNENRPS-----LKSLTQNRPEKSKCPLMGKXYFLDPSVTISEKLOKDKLGR 74  
 DB 396 VQNRGLPAPAVIILHLASTNIPYTSSEKSEVAI-----PELIDETRLA--LQVGR 448  
 0Y 75 VEEFLSKDISYLSLNKKEAKFAOTLGRISVPSPESAVTAETTSPPHSIDGSSFKSPDV 134  
 DB 449 LKEYLERKSRQKKRKE-----EMIGKVLPLAKK--VCEIUEKEP-----LEIDRI 495  
 0Y 135 CLSRGKLVEKAIKHDIFPSNSILSNALSMGWKILHIDIRYITQKKKELYLLKSSST 194







RA	Andresson S., Bennett J., Birch D.G., Fishman G.A., Hoffman D.R.,
RA	Inana G., Jacobson S.G., Musarella M.A., Steving P.A., Swarcop A.;
RT	"Spectrum of mutations in the RPGR gene that are identified in 20% of
RT	families with X-linked retinitis pigmentosa.";
RL	Am. J. Hum. Genet. 61:1287-1292(1997).
RN	[5]
RP	VARIANT RP3 VAL-60.
RX	MEDLINE-99070804; PubMed-9855162;
RA	Fishman G.A., Grover S., Jacobson S.G., Alexander K.R., Derlacki D.J.,
RA	Wu W., Buracynska M., Swaroop A.;
RT	X-linked retinitis pigmentosa in two families with a missense
RT	mutation in the RPGR gene and putative change of glycine to valine at
RT	codon 60.";
RL	Ophthalmology 105:2286-2296(1998).
CC	-1- FUNCTION: COULD BE GUANINE-NUCLEOTIDE RELEASING FACTOR.
CC	-1- SUBCELLULAR LOCATION: POSSIBLY MEMBRANE-ANCHORED.
CC	-1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM (SHOWN HERE) AND A
CC	SHORT FORM: ARE PRODUCED BY ALTERNATIVE SPLICING. THE SMALL FORM
CC	LACKS PART OF REPEAT 7..
CC	-1- TISSUE SPECIFICITY: HEART, BRAIN, PLACENTA, LUNG, LIVER, MUSCLE,
CC	KIDNEY, PANCREAS AND FETAL RETINAL PIGMENT EPITHELIUM.
CC	-1- DISEASE: DEFECTS IN RPGR ARE RESPONSIBLE FOR X-LINKED RETINITIS
CC	PIGMENTOSA-3 (XLRP-3 OR RP3), A FORM OF CHOROIO-RETINAL
CC	DGENERATION WHICH IS DISTINGUISHED FROM OTHER TYPES BY THE
CC	PRESENCE IN HETEROZYGOUS WOMEN OF A TAPEYAL-LIKE RETINAL REFLEX (A
CC	BRIGHTLY SCINTILLATING, GOLDEN-HUED, PATCHY APPEARANCE MOST
CC	STRIKING AROUND THE MACULA) BUT NO VISUAL DEFECT.
CC	-1- SIMILARITY: CONTAINS 7 RCL1 REPEATS.
CC	-1- DATABASE: NAME-Mutations of the RPGR gene:
CC	NOTE-Retina International's Scientific Newsletter;
CC	WWW-http://www.retina-international.com/sci_news/rpgmut.htm".
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/isb-ch).
CC	or send an email to license@lsb-sib.ch).
CC	-----
DR	EMBL: U57629; AAC50481.1; "-
DR	EMBL: X97668; CAA66258.1; "-
DR	MIM: 312610; "-
DR	InterPro: IPR000408; RCC1.
DR	Pfam: PF00415; RCC1; 7.
DR	PROSITE: PS00625; RCC1_1; FALSE_NEG.
DR	PROSITE: PS00626; RCC1_2; 4.
DR	PROSITE: PSS0012; RCC1_3; 6.
KW	Guanine-nucleotide releasing factor; Alternative splicing; Repeat;
KW	Vision; Retinitis pigmentosa; Disease mutation.
FT	REPEAT 11 52
FT	REPEAT 53 104
FT	REPEAT 106 157
FT	REPEAT 158 207
FT	REPEAT 208 260
FT	REPEAT 262 312
FT	REPEAT 314 366
FT	VARSPLIC 354 415 MISSING (IN SHORT ISOFORM).
FT	VARVARIANT 60 60 G-> V (IN RP3).
FT	VARIANT 60 60 G-> N (IN RP3).
FT	VARIANT 75 75 /FtId-VAR_008502.
FT	VARIANT 98 98 I-> V (IN RP3).
FT	VARIANT 130 130 H-> O (IN RP3).
FT	VARIANT 215 215 /FtId-VAR_008504.
FT	VARIANT 235 235 F-> C (IN RP3).
FT	VARIANT 250 250 /FtId-VAR_006850.
FT	VARIANT 260 260 G-> V (IN RP3).
FT	VARIANT 270 270 /FtId-VAR_008505.
FT	VARIANT 280 280 P-> S (IN RP3).
FT	VARIANT 290 290 /FtId-VAR_006851.
FT	VARIANT 300 300 C-> R (IN RP3).

FT	VARIANT	262		262	/FtId=VAR_008506.
FT	VARIANT	275		275	A -> G (IN RP3).
FT	VARIANT	275		275	/FtId=VAR_008507.
FT	VARIANT	425		425	G -> S (IN RP3).
FT	VARIANT	425		425	/FtId=VAR_006852.
FT	VARIANT	431		431	R -> K (IN RP3).
FT	VARIANT	431		431	/FtId=VAR_008508.
FT	VARIANT	436		436	I -> V (IN RP3).
FT	VARIANT	436		436	/FtId=VAR_008509.
FT	VARIANT	566		566	G -> D (IN RP3).
FT	VARIANT	566		566	/FtId=VAR_008510.
FT	VARIANT	566		566	G -> E (IN RP3).
FT	VARIANT	566		566	/FtId=VAR_008511.
FT	VARIANT	566		566	MRE -> MAKLRSTTAL (IN REF. 3).
FT	VARIANT	566		566	/FtId=VAR_008512.
FT	VARIANT	566		566	/FtId=VAR_008513.
FT	VARIANT	566		566	/FtId=VAR_008514.
FT	VARIANT	566		566	/FtId=VAR_008515.
FT	VARIANT	566		566	/FtId=VAR_008516.
FT	VARIANT	566		566	/FtId=VAR_008517.
FT	VARIANT	566		566	/FtId=VAR_008518.
FT	VARIANT	566		566	/FtId=VAR_008519.
FT	VARIANT	566		566	/FtId=VAR_008520.
FT	VARIANT	566		566	/FtId=VAR_008521.
FT	VARIANT	566		566	/FtId=VAR_008522.
FT	VARIANT	566		566	/FtId=VAR_008523.
FT	VARIANT	566		566	/FtId=VAR_008524.
FT	VARIANT	566		566	/FtId=VAR_008525.
FT	VARIANT	566		566	/FtId=VAR_008526.
FT	VARIANT	566		566	/FtId=VAR_008527.
FT	VARIANT	566		566	/FtId=VAR_008528.
FT	VARIANT	566		566	/FtId=VAR_008529.
FT	VARIANT	566		566	/FtId=VAR_008530.
FT	VARIANT	566		566	/FtId=VAR_008531.
FT	VARIANT	566		566	/FtId=VAR_008532.
FT	VARIANT	566		566	/FtId=VAR_008533.
FT	VARIANT	566		566	/FtId=VAR_008534.
FT	VARIANT	566		566	/FtId=VAR_008535.
FT	VARIANT	566		566	/FtId=VAR_008536.
FT	VARIANT	566		566	/FtId=VAR_008537.
FT	VARIANT	566		566	/FtId=VAR_008538.
FT	VARIANT	566		566	/FtId=VAR_008539.
FT	VARIANT	566		566	/FtId=VAR_008540.
FT	VARIANT	566		566	/FtId=VAR_008541.
FT	VARIANT	566		566	/FtId=VAR_008542.
FT	VARIANT	566		566	/FtId=VAR_008543.
FT	VARIANT	566		566	/FtId=VAR_008544.
FT	VARIANT	566		566	/FtId=VAR_008545.
FT	VARIANT	566		566	/FtId=VAR_008546.
FT	VARIANT	566		566	/FtId=VAR_008547.
FT	VARIANT	566		566	/FtId=VAR_008548.
FT	VARIANT	566		566	/FtId=VAR_008549.
FT	VARIANT	566		566	/FtId=VAR_008550.
FT	VARIANT	566		566	/FtId=VAR_008551.
FT	VARIANT	566		566	/FtId=VAR_008552.
FT	VARIANT	566		566	/FtId=VAR_008553.
FT	VARIANT	566		566	/FtId=VAR_008554.
FT	VARIANT	566		566	/FtId=VAR_008555.
FT	VARIANT	566		566	/FtId=VAR_008556.
FT	VARIANT	566		566	/FtId=VAR_008557.
FT	VARIANT	566		566	/FtId=VAR_008558.
FT	VARIANT	566		566	/FtId=VAR_008559.
FT	VARIANT	566		566	/FtId=VAR_008560.
FT	VARIANT	566		566	/FtId=VAR_008561.
FT	VARIANT	566		566	/FtId=VAR_008562.
FT	VARIANT	566		566	/FtId=VAR_008563.
FT	VARIANT	566		566	/FtId=VAR_008564.
FT	VARIANT	566		566	/FtId=VAR_008565.
FT	VARIANT	566		566	/FtId=VAR_008566.
FT	VARIANT	566		566	/FtId=VAR_008567.
FT	VARIANT	566		566	/FtId=VAR_008568.
FT	VARIANT	566		566	/FtId=VAR_008569.
FT	VARIANT	566		566	/FtId=VAR_008570.
FT	VARIANT	566		566	/FtId=VAR_008571.
FT	VARIANT	566		566	/FtId=VAR_008572.
FT	VARIANT	566		566	/FtId=VAR_008573.



```
FT METAL 1035 1035 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 1039 1039 COPPER (TYPE 1) (BY SIMILARITY).
FT METAL 1044 1044 COPPER (TYPE 1) (BY SIMILARITY).
FT CARBOHYD 138 138 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 226 226 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 396 396 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 582 582 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 756 756 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 920 920 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 271 271 G -> A (IN REF. 2).
FT CONFLICT 604 605 ED -> DN (IN REF. 2).
FT CONFLICT 823 823 T -> S (IN REF. 2).
FT CONFLICT 833 833 V -> L (IN REF. 2).
FT CONFLICT 868 868 C -> V (IN REF. 2).
FT CONFLICT 891 891 L -> R (IN REF. 2).
SQ SEQUENCE 1059 AA; 120840 MW; 12BA3B990A0B95E3 CRC64;
```

Query Match 7.0%; Score 84; DB 1; Length 1059;  
Best Local Similarity 20.8%; Pred. No. 44; Mismatches 9;  
Matches 44; Conservative 29; Indels 56; Gaps 9;

```
OY 10 SKGHPQGGIQVKNENKRPRLSKLKTNDNRPKSKCKPLW---GKVFYLDLPSTISEKIQ 65
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 52 SNFYLRNGPDRIGRKRYKKALYSEYDGFPTKTIDKPWMLGFLGPVL----KAEGDKVS 106
   : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 66 KDIKDLGGRVEEFLSKDISYLIISNKKKAKFAOTLGRISP-----VPSPEGAY 112
   : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 107 VHVKNFASRPYTFHAAGVYTKANE-----GAIYPDNTTDFORADPKLFPGGQYLY 157
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 113 TAETSPHP-----SH-----DGSSFKSPDTVCLSRGKILYEKAIR-DHDF 152
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 158 VLRANEPSPGEGDSNCVRIYHSHVDAPKDIASGLIGPLILC-KKGSILKKEKNIDOLF 216
   : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 153 IPSNSILSNALSWGKYLHIDDIRYIEQKKR 184
   : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 217 VLMFSVVDENLSWYLE---DNIKTFCSPEPK 244
```

Search completed: December 27, 2001, 16:56:15  
Job time: 230 sec

**THIS PAGE BLANK (USPTO)**